

# SUPER-RESOLUTION BY MEANS OF BEURLING MINIMAL EXTRAPOLATION

JOHN J. BENEDETTO AND WEILIN LI

**ABSTRACT.** Let  $M(\mathbb{T}^d)$  be the space of complex bounded Radon measures defined on the  $d$ -dimensional torus group  $(\mathbb{R}/\mathbb{Z})^d = \mathbb{T}^d$ , equipped with the total variation norm  $\|\cdot\|$ ; and let  $\hat{\mu}$  denote the Fourier transform of  $\mu \in M(\mathbb{T}^d)$ . We address the super-resolution problem: For given spectral (Fourier transform) data defined on a finite set  $\Lambda \subseteq \mathbb{Z}^d$ , determine if there is a unique  $\mu \in M(\mathbb{T}^d)$  of minimal norm for which  $\hat{\mu}$  equals this data on  $\Lambda$ . Without additional assumptions on  $\mu$  and  $\Lambda$ , our main theorem shows that the solutions to the super-resolution problem, which we call *minimal extrapolations*, depend crucially on the set  $\Gamma \subseteq \Lambda$ , defined in terms of  $\mu$  and  $\Lambda$ . For example, when  $\#\Gamma = 0$ , the minimal extrapolations are singular measures supported in the zero set of an analytic function, and when  $\#\Gamma \geq 2$ , the minimal extrapolations are singular measures supported in the intersection of  $\binom{\#\Gamma}{2}$  hyperplanes. By theory and example, we show that the case  $\#\Gamma = 1$  is different from other cases and is deeply connected with the existence of positive minimal extrapolations. This theorem has implications to the possibility *and* impossibility of uniquely recovering  $\mu$  from  $\Lambda$ . We illustrate how to apply our theory to both directions, by computing pertinent analytical examples. These examples are of interest in both super-resolution and deterministic compressed sensing. Our concept of an *admissibility range* fundamentally connects Beurling's theory of minimal extrapolation [Beu89a, Beu89b] with Candès and Fernandez-Granda's work on super-resolution [CFG14]. This connection is exploited to address situations where current algorithms fail to compute a numerical solution to the super-resolution problem.

## 1. INTRODUCTION

**1.1. Motivation.** The term *super-resolution* varies depending on the field, and, consequently, there are various types of super-resolution problems. In some situations [PPK03], super-resolution refers to the process of up-sampling an image onto a finer grid, which is a *spatial interpolation* procedure. In other situations [Lin12], super-resolution refers to the process of recovering the object's high frequency information from its low frequency information, which is a *spectral extrapolation* procedure. In both situations, the super-resolution problem is ill-posed because the missing information can be arbitrary. However, it is possible to provide meaningful super-resolution algorithms by using *prior knowledge* of the data. We develop a mathematical theory of the spectral extrapolation version, which we simply refer to as the *super-resolution problem*, see Problem (SR) for a precise statement.

To be concrete, our exposition focuses on imaging applications, but super-resolution ideas are of interest in other fields, e.g., [Rie99, KLM04]. In such applications, an image is obtained by convolving the object with the *point spread function* of an optical lens. Alternatively, the Fourier transform of the object is multiplied by a *modulation transfer function*. The resulting image's resolution is inherently limited by the Abbe diffraction

---

2010 *Mathematics Subject Classification.* 43A25, 46E27, 46N10, 94A08, 94A12.

*Key words and phrases.* Super-resolution, minimal extrapolation, total variation optimization.

limit, which depends on the illumination light's wavelength and on the diameter of the optical lens. Thus, the optical lens acts as a low-pass filter, see [Lin12]. The purpose of super-resolution is to use prior knowledge about the object to obtain an accurate image whose resolution is higher than what can be measured by the optical lens.

We mention two specific imaging problems, and they motivate the theory that we develop.

- (a) In astronomy [PK05], each star can be modeled as a complex number times a Dirac  $\delta$ -measure, and the Fourier transform of each star encodes important information about that star. However, an image of two stars that are close in distance resembles an image of a single star. In this context, the super-resolution problem is to determine the number of stars and their locations, using the prior information that the actual object is a linear combination of Dirac  $\delta$ -measures.
- (b) In medical imaging [Gre09], machines capture the structure of the patient's body tissues, in order to detect for anomalies in the patient. Their shapes and locations are the most important features, so each tissue can be modeled as the characteristic function of a closed set, or as a surface measure supported on the boundary of a set. The super-resolution problem is to capture the fine structures of the tissues, given that the actual object is a linear combination of singular measures.

**1.2. Problem statement.** We model objects as elements of  $M(\mathbb{T}^d)$ , the space of complex bounded Radon measures on  $\mathbb{T}^d = (\mathbb{R}/\mathbb{Z})^d$ , the  $d$ -dimensional torus group.  $M(\mathbb{T}^d)$  equipped with the total variation norm  $\|\cdot\|$  is a Banach algebra with unit, the Dirac  $\delta$ -measure, where multiplication is defined by convolution  $*$ . The *Fourier transform* of  $\mu \in M(\mathbb{T}^d)$  is the function  $\hat{\mu}: \mathbb{Z}^d \rightarrow \mathbb{C}$ , whose  $m$ -th Fourier coefficient is defined as

$$\hat{\mu}(m) = \int_{\mathbb{T}^d} e^{-2\pi i m \cdot x} d\mu(x).$$

See [BC09] for further details.

We consider the *super-resolution problem*: For a given finite subset  $\Lambda \subseteq \mathbb{Z}^d$  and given spectral data  $F$  of the form  $\hat{\mu}|_{\Lambda}$  for some unknown  $\mu \in M(\mathbb{T}^d)$ , find

$$(SR) \quad \arg \min_{\nu \in M(\mathbb{T}^d)} \|\nu\| \quad \text{subject to} \quad \hat{\nu} = \hat{\mu} \quad \text{on } \Lambda.$$

This is a *convex minimization problem* and we interpret a solution as a *simple* or *least complicated* high resolution extrapolation of  $F$ .

- Remark 1.1** (Image processing). (a) A primary objective is to determine if  $\mu \in M(\mathbb{T}^d)$  is a solution to the super-resolution problem given its spectral data  $\hat{\mu}|_{\Lambda}$ . The current literature has focused on discrete  $\mu \in M(\mathbb{T}^d)$ , e.g., see [CFG14, TBSR13, DCG12]. However, in the context of image processing,  $\mu$  is the unknown high resolution image, and a typical image cannot be approximated by a discrete measure. Hence, it is important to determine if non-discrete measures are solutions to the super-resolution problem. For this reason, we have formulated and shall study Problem (SR) in an abstract way.
- (b) Further, in the context of image processing, we can think of being given information that represents an image  $\mu \in M(\mathbb{T}^d)$ , and this information is in the form,

$$F(x) = (\mu * \psi)(x) = \int_{\mathbb{T}^d} \psi(x - y) d\mu(y),$$

for some  $\psi: \mathbb{T}^d \rightarrow \mathbb{C}$ . For simplicity, we assume that  $\widehat{\psi} = \mathbf{1}_\Lambda$ , the characteristic function of some finite set  $\Lambda \subseteq \mathbb{Z}^d$ . Then, we have

$$F = (\mu * \psi) = (\widehat{\mu} \mid_\Lambda)^\vee.$$

Thus, we interpret  $\widehat{\mu} \mid_\Lambda$  as the given low frequency information of  $\mu$ , and  $(\widehat{\mu} \mid_\Lambda)^\vee$  as the given low resolution image, even though in applications, we do not know the desired object  $\mu$ . This is the background for our formulation of Problem (SR).

With regard to Problem (SR), a fundamental question of uniqueness must be addressed. To see why this is so, if  $\mu$  is not the unique solution to the super-resolution problem, then the output of a numerical algorithm is not guaranteed to approximate  $\mu$ . Thus, it is important to determine sufficient conditions such that  $\mu$  is the unique solution. For this reason, we say that *super-resolution reconstruction of  $\mu$  from  $\Lambda$  is possible*, or *it is possible to super-resolve  $\mu$  from  $\Lambda$* , if and only if  $\mu$  is the unique solution to Problem (SR). Of course, it could be theoretically be possible to reconstruct  $\mu$  by other means.

**1.3. Background.** Problem (SR) was first studied by Candès and Fernandez-Granda [CFG14] and was inspired by results from compressed sensing [Don06, CRT06]. By considering different models of images and/or alternative measurement processes, it is possible to derive related (noiseless) “super-resolution” problems, e.g., see [Don92, DCG12, TBSR13, ASB14, ASB15]. There are several papers that address “super-resolution” in the presence of noise, e.g., see [CFG13, BTR13, ADCG15, TBR15, DP15].

We briefly discuss the important results of Candès and Fernandez-Granda [CFG14]. Given  $\Lambda = \{-M, -M+1, \dots, M\}^d \subseteq \mathbb{Z}^d$ , we say  $\mu = \sum_{k=1}^K a_k \delta_{x_k} \in M(\mathbb{T}^d)$  satisfies the *minimum separation condition with constant  $C > 0$*  if

$$(1.1) \quad \inf_{\substack{1 \leq j, k \leq K \\ j \neq k}} \|x_j - x_k\|_{\ell^\infty(\mathbb{T}^d)} \geq \frac{C}{M}.$$

**Theorem 1.2** (Candès and Fernandez-Granda, Theorems 1.2-1.3, [CFG14]. Fernandez-Granda, Theorem 2.2, [FG16]). *Let  $\mu = \sum_{k=1}^K a_k \delta_{x_k} \in M(\mathbb{T}^d)$  and  $\Lambda = \{-M, -M+1, \dots, M\}^d \subseteq \mathbb{Z}^d$ .*

- (a) *If  $d = 1$ ,  $M \geq 128$ , and  $\mu$  satisfies the minimum separation condition with  $C = 2$ , then  $\mu$  is the unique solution to Problem (SR). If additionally,  $\mu$  is real valued, then the constant can be reduced to  $C = 1.87$ .*
- (b) *If  $d = 1$ ,  $M \geq 10^3$ , and  $\mu$  satisfies the minimum separation condition with  $C = 1.26$ , then  $\mu$  is the unique solution to Problem (SR).*
- (c) *If  $d = 2$ ,  $\mu$  is real valued,  $M \geq 512$ , and  $\mu$  satisfies the minimum separation condition for  $C = 2.38$ , then  $\mu$  is the unique solution to Problem (SR).*

The theorem proves that, in theory, it is possible to uniquely recover discrete measures satisfying a sufficiently strong minimum separation condition. To compute a numerical approximation of such a measure, they proposed a convex-optimization based algorithm which relies on [Dum07, Theorem 4.24]. They stated their algorithm for  $d = 1$ , but it is possible an analogous algorithm for higher dimensions.

**Algorithm 1.3** (Candès and Fernandez-Granda, Section 4, [CFG14]). *Suppose  $\mu \in M(\mathbb{T})$  and  $\Lambda = \{-M, -M+1, \dots, M\} \subseteq \mathbb{Z}$  satisfy the hypotheses in any of the three situations described in Theorem 1.2.*

- (1) Solve the convex optimization problem described in [CFG14, Corollary 4.1] to obtain a function  $\varphi(x) = \sum_{m=-M}^M c_m e^{2\pi i m x}$  such that  $\|\varphi\|_\infty \leq 1$  and  $\langle \varphi, \mu \rangle = \|\mu\|$ . Then,  $\mu$  is supported in the set,

$$S = \{x \in \mathbb{T} : |\varphi(x)| = 1\}.$$

- (2) Assume that  $S \neq \mathbb{T}$ . Then,  $S$  is a discrete set that consists of at most  $2M$  points, and it is possible to determine  $\mu$  by solving a system of over-determined equations.

Candès and Fernandez-Granda pointed out that their algorithm for recovering  $\mu$  is incomplete because it fails precisely when  $\{x \in \mathbb{T} : |\varphi(x)| = 1\} = \mathbb{T}$ , i.e., when  $|\varphi| \equiv 1$ , since in this case,  $\varphi$  does not provide information about the support of  $\mu$ . We emphasize that this scenario can occur even if  $\mu \in M(\mathbb{T})$  and  $\Lambda \subseteq \mathbb{Z}$  satisfy the hypotheses of Theorem 1.2.

**1.4. Our approach.** This paper connects the Candès and Fernandez-Granda theory on super-resolution [CFG14] with Beurling's work on minimal extrapolation [Beu89a, Beu89b]. This connection is exploited to obtain new theoretical and computational results on super-resolution.

We first introduce some additional notation, since referring to Problem (SR) can be ambiguous when working with several different measures. We adopt Beurling's language, and we say that  $\nu \in M(\mathbb{T}^d)$  is a *minimal extrapolation* of  $\mu \in M(\mathbb{T}^d)$  from  $\Lambda \subseteq \mathbb{Z}^d$  if it is a solution to Problem (SR). If  $\nu \in M(\mathbb{T}^d)$  and  $\widehat{\mu} = \widehat{\nu}$  on  $\Lambda$ , then we say  $\nu$  is an *extrapolation* of  $\mu$  from  $\Lambda$ . When there is no ambiguity, we shall say that “ $\nu$  is an extrapolation (respectively, a minimal extrapolation)” as a shortened version of “ $\nu$  is an extrapolation (respectively, a minimal extrapolation) of  $\mu$  from  $\Lambda$ ”.

Let  $\mathcal{E} = \mathcal{E}(\mu, \Lambda)$  be the set of all minimal extrapolations. For any  $\nu \in \mathcal{E}$ , let  $\varepsilon = \varepsilon(\mu, \Lambda) = \|\nu\|$ , which is the optimal value attained by Problem (SR). It is possible to compute a numerical approximation of  $\varepsilon$  by solving the convex optimization problem described in [CFG14, Corollary 4.1], but its exact value is unknown. Finally, we define the set,

$$(1.2) \quad \Gamma = \Gamma(\mu, \Lambda) = \{m \in \Lambda : |\widehat{\mu}(m)| = \varepsilon(\mu, \Lambda)\}.$$

Although  $\Gamma$  is defined in terms of the unknown quantity  $\varepsilon$ , we shall see that, for many important applications, it is possible to deduce  $\varepsilon$  and  $\Gamma$ . Our theorem shows that the minimal extrapolations intimately depend on  $\Gamma$ .

**Theorem 1.4.** *Let  $\mu \in M(\mathbb{T}^d)$  and let  $\Lambda \subseteq \mathbb{Z}^d$  be a finite subset.*

- (a) *Suppose  $\#\Gamma = 0$ . Then, there exists a closed set  $S$  of  $d$ -dimensional Lebesgue measure zero such that each minimal extrapolation is a singular measure supported in  $S$ . In particular, if  $d = 1$ , then  $S$  is a finite number of points and each minimal extrapolation is a discrete measure supported in  $S$ .*
- (b) *Suppose  $\#\Gamma \geq 2$ . For each distinct pair  $m, n \in \Gamma$ , define  $\alpha_{m,n} \in \mathbb{R}/\mathbb{Z}$  by  $e^{2\pi i \alpha_{m,n}} = \widehat{\mu}(m)/\widehat{\mu}(n)$  and the closed set,*

$$(1.3) \quad S = \bigcap_{\substack{m,n \in \Gamma \\ m \neq n}} \{x \in \mathbb{T}^d : x \cdot (m - n) + \alpha_{m,n} \in \mathbb{Z}\},$$

*which is an intersection of  $\binom{\#\Gamma}{2}$  periodic hyperplanes. Then, each minimal extrapolation is a singular measure supported in  $S$ . In particular, if  $d = 1$ , then  $S$  is a finite number of points and each minimal extrapolation is a discrete measure supported in  $S$ .*

If  $d \geq 2$  and there exist  $d$  linearly independent vectors,  $p_1, p_2, \dots, p_d \in \mathbb{Z}^d$ , such that

$$\{p_1, p_2, \dots, p_d\} \subseteq \{m - n : m, n \in \Gamma\},$$

then  $S$  is a lattice on  $\mathbb{T}^d$ .

The following topics encapsulate the contributions of this paper.

- (a) *Qualitative behavior of minimal extrapolations.* It is interesting to note that  $\Gamma$  provides significant information about the minimal extrapolations. Theorem 1.4 shows that, regardless of the dimension, when  $\#\Gamma = 0$  or  $\#\Gamma \geq 2$ , the minimal extrapolations are always singular measures. However, when  $\#\Gamma = 1$ , the minimal extrapolations can be extremely different from each other. Proposition 2.8 shows that the case  $\#\Gamma = 1$  is connected with the existence of positive absolutely continuous minimal extrapolations. In Example 3.3, we provide an example for which there are uncountably many discrete minimal extrapolations as well as uncountably many positive absolutely continuous minimal extrapolations.
- (b) *Computational consequences.* One important aspect of Theorem 1.4 is its relationship with Algorithm 1.3. Proposition 2.6 shows that, if the algorithm fails, then we necessarily have  $\varepsilon = \|\hat{\mu}\|_{\ell^\infty(\Lambda)}$  and  $\Gamma \neq \emptyset$ . Since we are given the values of  $\hat{\mu}$  on  $\Lambda$ , this immediately tells us what  $\varepsilon$  and  $\Gamma$  are. If  $\#\Gamma = 1$ , then we cannot deduce anything about the minimal extrapolations. As previously discussed, this is inherent with the super-problem and is not an artifact of our analysis. If  $\#\Gamma \geq 2$ , then the theorem is applicable and the minimal extrapolations are singular measures supported in the set defined in (1.3), which can be explicitly computed. We show how to apply our theorem to compute pertinent analytical examples in Section 3. Hence, our theorem is applicable even when Algorithm 1.3 fails.
- (c) *Super-resolution of singular measures.* When  $d \geq 2$ , Theorem 1.4 suggests that some singular continuous measures could be solutions to the super-resolution problem. Somewhat surprisingly, this is indeed the case because Example 3.7 provides an example of a singular continuous minimal extrapolation. This also demonstrates that the conclusion of Theorem 1.4b is optimal. Determining what types of singular continuous measures can be uniquely recovered by Problem (SR) is an interesting mathematical problem, and we believe such a result would be useful in applications, e.g., [Gre09]. We shall carefully address the super-resolution of singular continuous measures in the sequel [BL16].
- (d) *Impossibility of super-resolution.* Theorem 1.4 does not require additional assumptions on  $\mu \in M(\mathbb{T}^d)$  or on the finite subset  $\Lambda \subseteq \mathbb{Z}^d$ . Since the theorem also describes the support set of the minimal extrapolations of  $\mu$  from  $\Lambda$ , it is useful for determining whether it is possible for a given  $\mu$  to be a minimal extrapolation. To illustrate this point, in Example 3.6, we provide a simple proof that shows, in general, a minimal separation condition is necessary to super-resolve a discrete measure.
- (e) *Discrete-continuous correspondence.* The super-resolution problem can be viewed as a continuous analogue of the discrete Fourier compressed sensing problem that was studied in [CRT06]. Let  $\mathcal{F}_N \in \mathbb{C}^{N \times N}$  be the DFT matrix,  $x \in \mathbb{C}^N$  be an unknown vector, and  $\Lambda \subseteq \{0, 1, \dots, N-1\}$ . The goal is to recover  $x$  from  $\mathcal{F}_N x|_\Lambda$  by solving

$$(CS) \quad \arg \min_{y \in \mathbb{C}^N} \|y\|_{\ell^1} \quad \text{subject to} \quad \mathcal{F}_N x = \mathcal{F}_N y \quad \text{on } \Lambda.$$

By considering the case that  $\mu = \sum_{n=1}^N a_n \delta_{n/N}$  and  $x(n) = a_n$ , it is straightforward to see that Problem (SR) is a generalization of Problem (CS), see [CFG14] for further

details. From this point of view, Theorem 1.4 is a discrete-continuous correspondence result. Indeed, if  $\#\Gamma \geq 2$ , and either  $d = 1$  or the vectors  $\{p_j\}$  exist, then the minimal extrapolations are necessarily discrete measures whose support lie on a lattice; as we just discussed, such measures can be identified with vectors that are solutions to the discrete problem.

- (f) *Mathematical connections.* Our results are closely related to Beurling's work on minimal extrapolation. He dealt with  $\mathbb{R}^1$  instead of  $\mathbb{T}^d$ , so Theorem 1.4 is an adaptation to the torus and a generalization to higher dimensions of [Beu89b, Theorem 2, page 362]. We remark that there are non-trivial technical differences between working with  $\mathbb{R}^1$  and  $\mathbb{T}^d$ . There are also other classical papers related to minimal extrapolation, e.g., [Ess45, Ess54, HR55].

Proposition 2.4 connects our results to the Candès and Fernandez-Granda theory [CFG14] by introducing a concept called an admissibility range for  $\varepsilon$ . This connection is exploited to prove Proposition 2.6, which in turn, shows that our theorem is applicable to situations where Algorithm 1.3 fails.

Proposition 2.1 contains basic duality results for Problem (SR) that can be derived from abstract convex analysis, e.g., see [DP15]. None of our proofs require knowledge of convex analysis, so we believe our paper is accessible to non-specialists.

**1.5. Outline.** Section 2 contains the basis of our mathematical theory. Section 2.1 proves well-known results on the super-resolution problem using only basic functional analysis. Section 2.2 provides a characterization of  $\Gamma$  and illustrates its importance in our analysis. The proof of Theorem 1.4 is given in Section 2.3. Section 2.4 contains the results that fully connect the Beurling and Candès and Fernandez-Granda theories, as well as the relationship between Theorem 1.4 and Algorithm 1.3. Section 2.5 includes a basic uniqueness result, and the mentioned non-uniqueness results for the case  $\#\Gamma = 1$ . Section 2.6 examines whether it is possible to relate the minimal extrapolations of  $\mu$  with those of  $T\mu$  for various operators  $T$ . Section 3 contains all of the examples that we have previously mentioned, including the minimal extrapolations of discrete measures and of singular continuous measures.

## 2. MATHEMATICAL THEORY

**2.1. Preliminary results.** Let  $C(\mathbb{T}^d)$  be the space of complex-valued continuous functions on  $\mathbb{T}^d$  equipped with the sup-norm  $\|\cdot\|_\infty$ . Then,  $C(\mathbb{T}^d)$  is a Banach space, and let  $C(\mathbb{T}^d)'$  be its dual space of continuous linear functionals with the usual norm  $\|\cdot\|$ . The celebrated Riesz representation theorem, e.g., see [BC09, Theorem 7.2.7, page 334] states that:

- (a) For each  $\mu \in M(\mathbb{T}^d)$ , there exists a bounded linear functional  $\ell_\mu \in C(\mathbb{T}^d)'$  such that  $\|\mu\| = \|\ell_\mu\|$  and

$$\forall f \in C(\mathbb{T}^d), \quad \ell_\mu(f) = \langle f, \mu \rangle = \int_{\mathbb{T}^d} f(x) \overline{d\mu(x)}.$$

- (b) For each bounded linear functional  $\ell \in C(\mathbb{T}^d)'$ , there exists a unique  $\mu \in M(\mathbb{T}^d)$  such that  $\|\mu\| = \|\ell\|$  and

$$\forall f \in C(\mathbb{T}^d), \quad \ell(f) = \langle f, \mu \rangle = \int_{\mathbb{T}^d} f(x) \overline{d\mu(x)}.$$

Proposition 2.1a shows that the super-resolution problem is well-posed, by using standard functional analysis arguments. However, this type of argument does not yield useful statements about the minimal extrapolations. Instead of working with  $C(\mathbb{T}^d)$ , we shall work with a subspace. Since only  $\widehat{\mu}|_\Lambda$  is important to the super-resolution problem, we consider the subspace

$$C(\mathbb{T}^d; \Lambda) = \left\{ f \in C(\mathbb{T}^d) : f(x) = \sum_{m \in \Lambda} a_m e^{2\pi i m \cdot x}, a_m \in \mathbb{C} \right\}.$$

Proposition 2.1b shows that  $C(\mathbb{T}^d; \Lambda)$  is a closed subspace of  $C(\mathbb{T}^d)$ , and that implies  $C(\mathbb{T}^d; \Lambda)$  is a Banach space. Further, Proposition 2.1c demonstrates that

$$U = U(\mathbb{T}^d; \Lambda) = \{f \in C(\mathbb{T}^d; \Lambda) : \|f\|_\infty \leq 1\},$$

the closed unit ball of  $C(\mathbb{T}^d; \Lambda)$ , is compact.

The purpose of restricting to the subspace,  $C(\mathbb{T}^d; \Lambda)$ , is to identify  $\mu \in M(\mathbb{T}^d)$  with the bounded linear functional,  $L_\mu \in C(\mathbb{T}^d; \Lambda)'$ , defined as

$$\forall f \in C(\mathbb{T}^d; \Lambda), \quad L_\mu(f) = \int_{\mathbb{T}^d} f(x) \overline{d\mu(x)}.$$

Although, by definition,  $\|L_\mu\| = \sup_{f \in U} |L_\mu(f)|$ , Proposition 2.1d,e show that we have the stronger statement,

$$\|L_\mu\| = \max_{f \in U} |L_\mu(f)| = \varepsilon.$$

The purpose of studying  $L_\mu$  is to deduce information about the minimal extrapolations. As a consequence of the Radon-Nikodym theorem, for each  $\mu \in M(\mathbb{T}^d)$ , there exists a  $\mu$ -measurable function  $\text{sign}(\mu)$  such that  $|\text{sign}(\mu)| = 1$   $\mu$ -a.e. and satisfies the identity

$$\forall f \in L^1_{|\mu|}(\mathbb{T}^d), \quad \int_{\mathbb{T}^d} f d|\mu| = \int_{\mathbb{T}^d} f \overline{\text{sign}(\mu)} d\mu.$$

See [BC09, Theorem 5.3.2, page 242, and Theorem 5.3.5, page 244] for further details. The support of  $\mu \in M(\mathbb{T}^d)$ , denoted  $\text{supp}(\mu)$ , is the complement of all open sets  $A \subseteq \mathbb{T}^d$  such that  $\mu(A) = 0$ . Proposition 2.1g shows that it is possible to interpolate the sign of the minimal extrapolations by a function belonging to  $U$ .

**Proposition 2.1.** *Let  $\mu \in M(\mathbb{T}^d)$  and let  $\Lambda \subseteq \mathbb{Z}^d$  be a finite subset.*

- (a)  $\mathcal{E} \subseteq M(\mathbb{T}^d)$  is non-empty, weak-\* compact, and convex.
- (b)  $C(\mathbb{T}^d; \Lambda)$  is a closed subspace of  $C(\mathbb{T}^d)$ .
- (c)  $U$  is a compact subset of  $C(\mathbb{T}^d; \Lambda)$ .
- (d)  $\varepsilon = \|L_\mu\|$ .
- (e)  $\varepsilon = \max_{f \in U} |\langle f, \mu \rangle|$ .
- (f) There exists  $\varphi \in U$  such that  $\langle \varphi, \mu \rangle = \varepsilon$ .
- (g) If  $\varphi \in U$  and  $\langle \varphi, \mu \rangle = \varepsilon$ , then

$$\forall \nu \in \mathcal{E}, \quad \varphi = \text{sign}(\nu) \quad \nu\text{-a.e.}, \quad \text{and} \quad \text{supp}(\nu) \subseteq \{x \in \mathbb{T}^d : |\varphi(x)| = 1\}.$$

*Proof.*

- (a) By definition of  $\varepsilon$ , there exists a sequence  $\{\nu_j\} \subseteq \{\nu : \widehat{\nu} = \widehat{\mu} \text{ on } \Lambda\}$  such that  $\|\nu_j\| \rightarrow \varepsilon$ . Then, this sequence is bounded. By Banach-Alaoglu, after passing to a subsequence, we can assume there exists  $\nu \in M(\mathbb{T}^d)$  such that  $\nu_j \rightarrow \nu$  in the weak-\* topology.

Let  $V$  be the closed unit ball of  $C(\mathbb{T}^d)$ . We have  $\|\nu\| \leq \varepsilon$  because

$$\|\nu\| = \sup_{f \in V} |\langle f, \nu \rangle| = \sup_{f \in V} \lim_{j \rightarrow \infty} |\langle f, \nu_j \rangle| \leq \sup_{f \in V} \lim_{j \rightarrow \infty} \|f\|_\infty \|\nu_j\| \leq \varepsilon.$$

Moreover, for  $m \in \Lambda$ , we have

$$\widehat{\mu}(m) = \lim_{j \rightarrow \infty} \widehat{\nu}_j(m) = \lim_{j \rightarrow \infty} \int_{\mathbb{T}^d} e^{-2\pi i m \cdot x} d\nu_j(x) = \int_{\mathbb{T}^d} e^{-2\pi i m \cdot x} d\nu(x) = \widehat{\nu}(m).$$

This shows that  $\nu$  is an extrapolation, and thus,  $\|\nu\| \geq \varepsilon$ . Therefore,  $\nu \in \mathcal{E}$ .

The proof that  $\mathcal{E}$  is weak-\* compact is similar. Pick any sequence  $\{\nu_j\} \subseteq \mathcal{E}$  and after passing to a subsequence, we can assume  $\nu_j \rightarrow \nu$  in the weak-\* topology for some  $\nu \in M(\mathbb{T}^d)$ . By the same argument, we see that  $\nu \in \mathcal{E}$ .

If  $\mathcal{E}$  contains exactly one measure, then  $\mathcal{E}$  is trivially convex. Otherwise, let  $t \in [0, 1]$ ,  $\nu_0, \nu_1 \in \mathcal{E}$ , and  $\nu_t = (1-t)\nu_0 + t\nu_1$ . Then,  $\nu_t$  is an extrapolation and thus,  $\|\nu_t\| \geq \varepsilon$ . By the triangle inequality, we have  $\|\nu_t\| \leq (1-t)\|\nu_0\| + t\|\nu_1\| = \varepsilon$ . Thus,  $\nu_t \in \mathcal{E}$  for each  $t \in [0, 1]$ .

- (b) Suppose  $\{f_j\} \subseteq C(\mathbb{T}^d)$  and that there exists  $f \in C(\mathbb{T}^d)$  such that  $f_j \rightarrow f$  uniformly. Then,  $\widehat{f}_j(m) \rightarrow \widehat{f}(m)$  for all  $m \in \mathbb{Z}^d$ . Since  $\widehat{f}_j(m) = 0$  if  $m \notin \Lambda$ , we deduce that  $\widehat{f}(m) = 0$  if  $m \notin \Lambda$ . This shows that  $f \in C(\mathbb{T}^d; \Lambda)$  and thus,  $C(\mathbb{T}^d; \Lambda)$  is a closed subspace of  $C(\mathbb{T}^d)$ .
- (c) Let  $\{f_j\} \subseteq U$ . We first show that  $\{f_j\}$  is a uniformly bounded equicontinuous family. By definition,  $\|f_j\|_\infty \leq 1$ , and there exist  $a_{j,m} \in \mathbb{C}$  such that  $f_j(x) = \sum_{m \in \Lambda} a_{j,m} e^{2\pi i m \cdot x}$ . Note that  $|a_{j,m}| \leq \|f_j\|_\infty \leq 1$ . Then, for any  $x, y \in \mathbb{T}^d$ , we have

$$|f_j(x) - f_j(y)| = \left| \sum_{m \in \Lambda} a_{j,m} (e^{2\pi i m \cdot x} - e^{2\pi i m \cdot y}) \right| \leq \sum_{m \in \Lambda} |e^{2\pi i m \cdot x} - e^{2\pi i m \cdot y}|.$$

Let  $\varepsilon > 0$  and  $m \in \Lambda$ . There exists  $\delta_m > 0$  such that  $|e^{2\pi i m \cdot x} - e^{2\pi i m \cdot y}| < \varepsilon$  whenever  $|x - y| < \delta_m$ . Let  $\delta = \min_{m \in \Lambda} \delta_m$ . Combining this with the previous inequalities, we have

$$|f_j(x) - f_j(y)| \leq \sum_{m \in \Lambda} |e^{2\pi i m \cdot x} - e^{2\pi i m \cdot y}| \leq \#\Lambda \varepsilon,$$

whenever  $|x - y| < \delta$ . This shows that  $\{f_j\}$  is a uniformly bounded equicontinuous family.

By the Arzelà-Ascoli theorem, there exists  $f \in C(\mathbb{T}^d)$ , with  $\|f\|_\infty \leq 1$ , and a subsequence  $\{f_{j_k}\}$  such that  $f_{j_k} \rightarrow f$  uniformly. Thus,  $\widehat{f}_{j_k}(m) \rightarrow \widehat{f}(m)$  for all  $m \in \mathbb{Z}^d$ , which shows that  $f \in C(\mathbb{T}^d; \Lambda)$ .

- (d) Let  $\nu \in \mathcal{E}$ . Then,

$$\forall f \in U, \quad |L_\mu(f)| = |\langle f, \mu \rangle| = |\langle f, \nu \rangle| \leq \|f\|_\infty \|\nu\| \leq \varepsilon,$$

which proves the upper bound,  $\|L_\mu\| \leq \varepsilon$ .

For the lower bound, we use the Hahn-Banach theorem to extend  $L_\mu \in C(\mathbb{T}^d; \Lambda)'$  to  $\ell \in C(\mathbb{T}^d)'$ , where  $\|L_\mu\| = \|\ell\|$ . By the Riesz representation theorem, there exists a unique  $\sigma \in M(\mathbb{T}^d)$  such that  $\ell(f) = \langle f, \sigma \rangle$  for all  $f \in C(\mathbb{T}^d)$  and  $\|\sigma\| = \|\ell\|$ . In particular,

$$\forall f \in C(\mathbb{T}^d; \Lambda), \quad \langle f, \sigma \rangle = \ell(f) = L_\mu(f) = \langle f, \mu \rangle.$$



Set  $f(x) = e^{-2\pi i m \cdot x}$ , where  $m \in \Lambda$ , to deduce that  $\hat{\mu} = \hat{\sigma}$  on  $\Lambda$ . This implies  $\|\sigma\| \geq \varepsilon$ . Combining these facts, we have

$$\varepsilon \leq \|\sigma\| = \|\ell\| = \|L_\mu\|.$$

This proves the lower bound.

- (e) We know that  $\|L_\mu\| = \varepsilon$ . By definition, there exists  $\{f_j\} \subseteq U$  such that  $|L_\mu(f_j)| \geq \varepsilon - 1/j$ . By compactness of  $U$ , there exists a subsequence  $\{f_{j_k}\} \subseteq U$  and  $f \in U$  such that  $f_{j_k} \rightarrow f$  uniformly. We immediately have  $|L_\mu(f)| \leq \|f\|_\infty \varepsilon \leq \varepsilon$ . For the reverse inequality, as  $k \rightarrow \infty$ ,

$$|L_\mu(f)| \geq |L_\mu(f_{j_k})| - |L_\mu(f - f_{j_k})| \geq \varepsilon - \frac{1}{j_k} - \|\mu\| \|f - f_{j_k}\|_\infty \rightarrow \varepsilon.$$

This proves that  $|L_\mu(f)| = \varepsilon$ .

- (f) There exists  $f \in U$  such that  $|L_\mu(f)| = \varepsilon$ . This implies  $e^{i\theta} L_\mu(f) = \varepsilon$  for some  $\theta \in \mathbb{R}$ . Hence, let  $\varphi = e^{i\theta} f \in U$ , and thus,  $L_\mu(\varphi) = \varepsilon$ .
- (g) Let  $\nu \in \mathcal{E}$ . Since  $\varphi \in U \subseteq C(\mathbb{T}^d; \Lambda)$ , we have

$$\langle \varphi, \nu \rangle = \langle \varphi, \mu \rangle = L_\mu(\varphi) = \varepsilon = \|\nu\|.$$

Since  $\|\varphi\|_\infty \leq 1$  and  $\langle \varphi, \nu \rangle = \|\nu\|$ , we must have  $\varphi = \text{sign}(\nu)$   $\nu$ -a.e.

Since  $|\varphi| = |\text{sign}(\nu)| = 1$   $\nu$ -a.e. and  $\nu$  is a Radon measure,

$$\text{supp}(\nu) \subseteq \overline{\{x \in \mathbb{T}^d : |\varphi(x)| = 1\}} = \{x \in \mathbb{T}^d : |\varphi(x)| = 1\}.$$

The last equality holds because the inverse image of the closed set  $\{1\}$  under the continuous function  $|\varphi|$  is closed.

□

**2.2. A characterization of  $\Gamma$ .** Proposition 2.1f,g show that there exists  $\varphi \in U$  such that the minimal extrapolations are supported in the closed set  $S = \{x \in \mathbb{T}^d : |\varphi(x)| = 1\}$ . By itself, this statement is not useful because if  $|\varphi| \equiv 1$ , then  $S = \mathbb{T}^d$ . However, Proposition 2.2 characterizes the case that  $|\varphi| \equiv 1$  in terms of the set  $\Gamma$ . Intuitively,  $\#\Gamma$  represents the number of “bad” functions in  $U$  that interpolate the signs of the minimal extrapolations. While it is desirable to have  $\Gamma = \emptyset$ , perhaps surprisingly, we can make strong statements when  $\#\Gamma$  is large.

**Proposition 2.2.** *Let  $\mu \in M(\mathbb{T}^d)$  and  $\Lambda \subseteq \mathbb{Z}^d$  be a finite subset.*

- (a) *Suppose  $\varphi \in U$ ,  $\langle \varphi, \mu \rangle = \varepsilon$ , and  $|\varphi| \equiv 1$ . Then,  $\Gamma \neq \emptyset$ .*
- (b) *For each  $m \in \mathbb{Z}^d$ , define  $\alpha_m \in \mathbb{R}/\mathbb{Z}$  by the formula  $e^{-2\pi i \alpha_m} \hat{\mu}(m) = |\hat{\mu}(m)|$ . If  $m \in \Gamma$ , then*

$$\forall \nu \in \mathcal{E}, \quad \text{sign}(\nu)(x) = e^{2\pi i \alpha_m} e^{2\pi i m \cdot x} \quad \nu\text{-a.e.}$$

*Proof.*

- (a) Since  $\varphi \in U$  and  $|\varphi| \equiv 1$ , we must have  $\varphi = e^{2\pi i \theta} e^{2\pi i m \cdot x}$  for some  $m \in \Lambda$  and  $\theta \in \mathbb{R}$ . Then, we have

$$\varepsilon = \langle \varphi, \mu \rangle = e^{-2\pi i \theta} \overline{\hat{\mu}(m)},$$

which shows that  $m \in \Gamma$ .

(b) Suppose  $m \in \Gamma$  and  $\nu \in \mathcal{E}$ . Then

$$\int_{\mathbb{T}^d} e^{-2\pi i \alpha_m} e^{-2\pi i m \cdot x} d\nu(x) = e^{-2\pi i \alpha_m} \widehat{\nu}(m) = e^{-2\pi i \alpha_m} \widehat{\mu}(m) = |\widehat{\mu}(m)| = \varepsilon = \|\nu\|.$$

This shows  $\text{sign}(\nu)(x) = e^{2\pi i \alpha_m} e^{2\pi i m \cdot x}$   $\nu$ -a.e.

□

**2.3. An analogue of Beurling's theorem.** We are ready to prove Theorem 1.4.

*Proof.*

(a) By Proposition 2.1, there exists  $\varphi \in U$  such that  $\langle \varphi, \mu \rangle = \varepsilon$ , and each minimal extrapolation is supported in the closed set

$$S = \{x \in \mathbb{T}^d : |\varphi(x)| = 1\}.$$

Note that  $|\varphi| \not\equiv 1$ . In fact, if  $|\varphi| \equiv 1$ , then Proposition 2.2a implies  $\Gamma \neq \emptyset$ , which is a contradiction. We have  $\varphi(x) = \sum_{m \in \Lambda} a_m e^{2\pi i m \cdot x}$  for some  $a_m \in \mathbb{C}$ . Consider the function

$$(2.1) \quad \Phi(x) = 1 - |\varphi(x)|^2 = 1 - \sum_{m \in \Lambda} \sum_{n \in \Lambda} a_m \overline{a_n} e^{2\pi i (m-n) \cdot x}.$$

Then, the minimal extrapolations are supported in the closed set

$$S = \{x \in \mathbb{T}^d : \Phi(x) = 0\}.$$

Note that  $\Phi \not\equiv 0$  because  $|\varphi| \not\equiv 1$ . Since  $\Phi$  is a non-trivial real-analytic function,  $S$  is a set of  $d$ -dimensional Lebesgue measure zero. In particular, if  $d = 1$ , then  $S$  is a finite set of points.

(b) Let  $m, n \in \Gamma$ . There exist  $\alpha_m, \alpha_n \in \mathbb{R}/\mathbb{Z}$  defined in Proposition 2.2b, such that

$$\forall \nu \in \mathcal{E}, \quad \text{sign}(\nu)(x) = e^{2\pi i \alpha_m} e^{2\pi i m \cdot x} = e^{2\pi i \alpha_n} e^{2\pi i n \cdot x} \quad \nu\text{-a.e.}$$

Set  $\alpha_{m,n} = \alpha_m - \alpha_n \in \mathbb{R}/\mathbb{Z}$ . Then each minimal extrapolation is supported in

$$S_{m,n} = \{x \in \mathbb{T}^d : x \cdot (m - n) + \alpha_{m,n} \in \mathbb{Z}\}.$$

Thus, *each minimal extrapolation is supported in the set*

$$S = \bigcap_{\substack{m,n \in \Gamma \\ m \neq n}} S_{m,n} = \bigcap_{\substack{m,n \in \Gamma \\ m \neq n}} \{x \in \mathbb{T}^d : x \cdot (m - n) + \alpha_{m,n} \in \mathbb{Z}\}.$$

Note that  $e^{2\pi i \alpha_{m,n}} = \widehat{\mu}(m)/\widehat{\mu}(n)$  because

$$e^{-2\pi i \alpha_m} \widehat{\mu}(m) = |\widehat{\mu}(m)| = \varepsilon = |\widehat{\mu}(n)| = e^{-2\pi i \alpha_n} \widehat{\mu}(n).$$

Suppose  $\{p_1, \dots, p_d\}$  satisfies the hypothesis. By the support assertion that we just proved, there exists  $\beta = (\beta_1, \beta_2, \dots, \beta_d) \in \mathbb{T}^d$  such that every minimal extrapolation is supported in

$$S = \bigcap_{j=1}^d \{x \in \mathbb{T}^d : x \cdot p_j + \beta_j \in \mathbb{Z}\}.$$

Let us explain the geometry of the situation before we proceed with the proof that  $S$  is a lattice. Note that  $\{x \in \mathbb{T}^d : x \cdot p_j + \beta_j \in \mathbb{Z}\}$  is a family of parallel and periodically spaced hyperplanes. Since the vectors,  $p_1, p_2, \dots, p_d$ , are assumed to be linearly independent, one family of hyperplanes is not parallel to any other family of hyperplanes. Hence,

the intersection of  $d$  non-parallel and periodically spaced hyperplanes is a lattice, see Figure 2.1 for an illustration.

For the rigorous proof of this observation, first note that

$$S = \{x \in \mathbb{T}^d : Px + \beta \in \mathbb{Z}^d\},$$

where  $P = (p_1, p_2, \dots, p_d)^t \in \mathbb{Z}^{d \times d}$  is invertible because its rows are linearly independent. Let  $x_0 \in \mathbb{R}^d$  be the solution to  $Px + \beta = 0$ , and let  $q_j \in \mathbb{Q}^d$  be the solution to  $Px = e_j$ , where  $e_j$  is the standard basis vector for  $\mathbb{R}^d$ . Then  $p_j \cdot q_k = \delta_{j,k}$ , and  $S$  is generated by the point  $x_0$  and the lattice vectors,  $q_1, q_2, \dots, q_d$ .

□

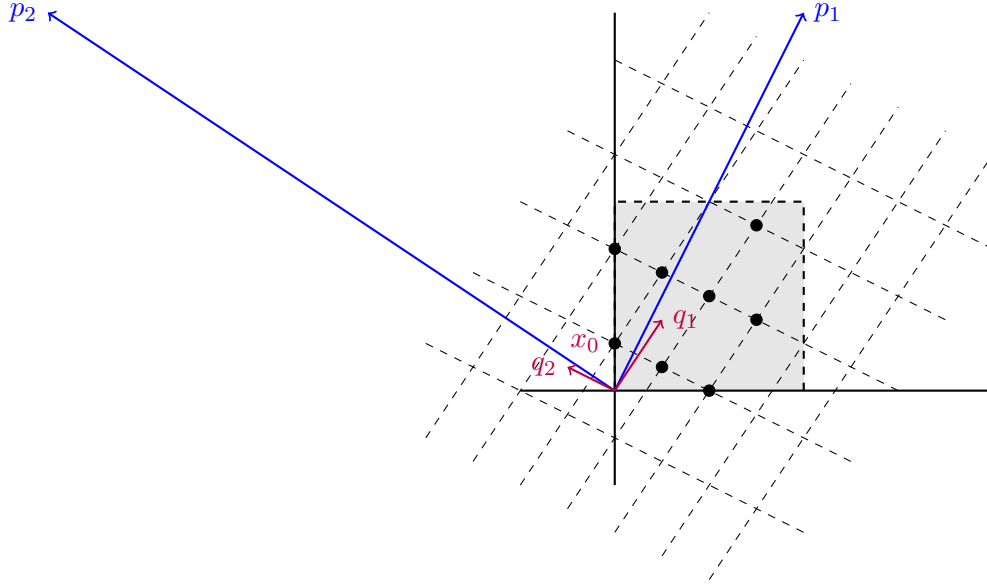


FIGURE 2.1. An illustration of Theorem 1.4b, where  $d = 2$ ,  $p_1 = (1, 2)$ ,  $p_2 = (-3, 2)$ ,  $\beta_1 = 1/2$ ,  $\beta_2 = -1/2$ ,  $q_1 = (1/4, 3/8)$ , and  $q_2 = (-1/4, 1/8)$ . The family of hyperplanes are the dashed lines, the lattice  $S$  is the black dots, and the shaded region is  $[0, 1]^2$ .

**Remark 2.3.** Example 3.7 provides an example where  $\#\Gamma = 3$  and there exists a singular continuous minimal extrapolation. This demonstrates that the conclusion of Theorem 1.4b is optimal.

**2.4. The admissibility range for  $\varepsilon$ .** While the mathematical theory we have developed connects the set  $\Gamma$  with the support of the minimal extrapolations, the difficulty of applying the theory is that in general,  $\varepsilon$  is unknown. However, in many important situations, it is possible to deduce the value of  $\varepsilon$ . We say  $[A, B] \subseteq \mathbb{R}^+$  is an *admissibility range* for  $\varepsilon$  provided that  $0 \leq A \leq \varepsilon \leq B$ . The following proposition shows that we have  $A = \|\widehat{\mu}\|_{\ell^\infty(\Lambda)}$  and  $B = \|\mu\|$  as an admissibility range, and  $B$  can be improved in certain situations.

**Proposition 2.4.** *Let  $\mu \in M(\mathbb{T}^d)$  and let  $\Lambda \subseteq \mathbb{Z}^d$  be a finite subset. We have the lower and upper bounds,*

$$(2.2) \quad \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} \leq \varepsilon \leq \|\mu\|.$$

Further, if there exists an extrapolation  $\nu \in M(\mathbb{T}^d)$  and  $\|\nu\| < \|\mu\|$ , then

$$(2.3) \quad \|\hat{\mu}\|_{\ell^\infty(\Lambda)} \leq \varepsilon \leq \|\nu\| < \|\mu\|.$$

*Proof.* To see the lower bound for  $\varepsilon$  in (2.2) and (2.3), let  $\sigma$  be a minimal extrapolation. Then,

$$\sup_{m \in \Lambda} |\hat{\mu}(m)| = \sup_{m \in \Lambda} |\hat{\sigma}(m)| \leq \|\sigma\| = \varepsilon.$$

The upper bounds,  $\varepsilon \leq \|\mu\|$  and  $\varepsilon \leq \|\nu\|$  in (2.2) and (2.3), follow by definition of  $\varepsilon$ .  $\square$

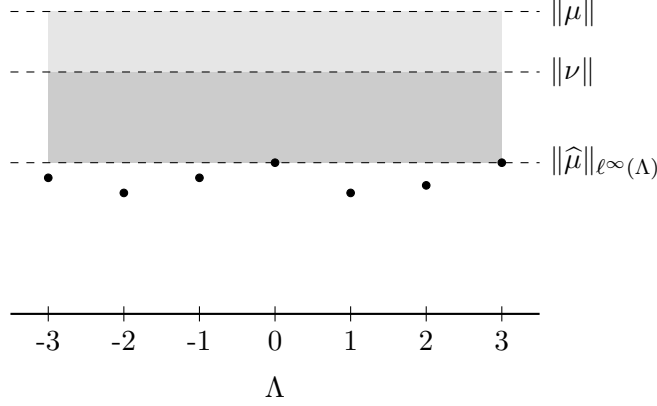


FIGURE 2.2. The black points represent the values of  $|\hat{\mu}|$  on  $\Lambda$ . The union of the light and dark gray regions is the admissibility range  $[\|\hat{\mu}\|_{\ell^\infty(\Lambda)}, \|\mu\|]$ , whereas the dark gray region is an improved admissibility range  $[\|\hat{\mu}\|_{\ell^\infty(\Lambda)}, \|\nu\|]$ .

**Remark 2.5.** By tightening the admissibility range for  $\varepsilon$ , we can deduce information about the minimal extrapolations. The simplest case is when  $\varepsilon = \|\hat{\mu}\|_{\ell^\infty(\Lambda)} = \|\mu\|$ , see Examples 3.2 and 3.3. The next simplest case is when there exists an extrapolation  $\nu$ , such that  $\varepsilon = \|\hat{\mu}\|_{\ell^\infty(\Lambda)} = \|\nu\|$ , see Example 3.4. A more complicated case is when  $\|\hat{\mu}\|_{\ell^\infty(\Lambda)} < \varepsilon$ , see Example 3.5.

Candès and Fernandez-Granda [CFG14] focused entirely on the case that  $\|\mu\| = \varepsilon$  because this is a necessary condition to super-resolve  $\mu$ . In contrast to their result, our theory is better suited for situations when  $\|\hat{\mu}\|_{\ell^\infty} = \varepsilon$ . This is because Theorem 1.4b is strongest when  $|\hat{\mu}(m)| = \varepsilon$  for many points  $m \in \Lambda$ , i.e., when  $\#\Gamma$  is large. Hence, our theory is better suited for deducing the impossibility of super-resolution reconstruction. Remarkably though, the following proposition shows that if Algorithm 1.3 fails, then  $\varepsilon = \|\hat{\mu}\|_{\ell^\infty}$ , and since we are given  $\hat{\mu}|_\Lambda$ , we immediately get the exact value of  $\varepsilon$ . Recall that the algorithm fails if the function  $\varphi \in U$  computed in the first step of the algorithm satisfies  $\langle \varphi, \mu \rangle = \varepsilon$  and  $|\varphi| \equiv 1$ .

**Proposition 2.6.** *Let  $\mu \in M(\mathbb{T}^d)$  and let  $\Lambda \subseteq \mathbb{Z}^d$  be a finite subset. Suppose there exists  $\varphi \in U$  such that  $|\varphi| \equiv 1$  and  $\langle \varphi, \mu \rangle = \varepsilon$ . Then,  $\Gamma \neq \emptyset$  and  $\varepsilon = \|\hat{\mu}\|_{\ell^\infty(\Lambda)}$ .*

*Proof.* By Proposition 2.2, we must have  $\Gamma \neq \emptyset$ . Let  $m \in \Gamma$  and by definition of  $\Gamma$ , we have  $|\hat{\mu}(m)| = \varepsilon$ . This combined with the lower bound in Proposition 2.4 proves that  $\varepsilon = \|\hat{\mu}\|_{\ell^\infty(\Lambda)}$ .  $\square$

**2.5. On uniqueness and non-uniqueness of minimal extrapolations.** Theorem 1.4 provides sufficient conditions for the minimal extrapolations to be supported in a discrete set, but it does not provide sufficient conditions for uniqueness. Since a family of discrete measures supported on a common set behaves essentially like a vector, we use basic linear algebra to address the question of uniqueness when the minimal extrapolations are supported in a discrete set.

**Proposition 2.7.** *Let  $\mu \in M(\mathbb{T}^d)$  and let  $\Lambda = \{m_1, m_2, \dots, m_J\} \subseteq \mathbb{Z}^d$  be a finite subset. Suppose there exists a finite set,  $\{x_k \in \mathbb{T}^d : k = 1, 2, \dots, K\}$ , such that each minimal extrapolation is supported in this set. Suppose that the matrix,*

$$(2.4) \quad E(m_1, \dots, m_J; x_1, \dots, x_K) = \begin{pmatrix} e^{-2\pi i m_1 \cdot x_1} & e^{-2\pi i m_1 \cdot x_2} & \dots & e^{-2\pi i m_1 \cdot x_K} \\ e^{-2\pi i m_2 \cdot x_1} & e^{-2\pi i m_2 \cdot x_2} & \dots & e^{-2\pi i m_2 \cdot x_K} \\ \vdots & \vdots & \dots & \vdots \\ e^{-2\pi i m_J \cdot x_1} & e^{-2\pi i m_J \cdot x_2} & \dots & e^{-2\pi i m_J \cdot x_K} \end{pmatrix},$$

*has full column rank (this can only occur if  $J \geq K$ ). Then, the minimal extrapolation is unique.*

*Proof.* Let  $\nu$  be the difference of any two minimal extrapolations. Then  $\nu$  is also supported in  $\{x_k : k = 1, 2, \dots, K\}$  and it is of the form  $\nu = \sum_{k=1}^K a_k \delta_{x_k}$ . Since  $\widehat{\nu} = 0$  on  $\Lambda$ , we have  $0 = \widehat{\nu}(m_j) = \sum_{k=1}^K a_k e^{-2\pi i m_j \cdot x_k}$  for  $j = 1, 2, \dots, J$ , which is equivalent to the linear system  $Ea = 0$ , where  $a = (a_1, \dots, a_K) \in \mathbb{C}^K$ . By assumption,  $E$  has full column rank. This implies  $a = 0$ .  $\square$

To finish this subsection, we address the situation when  $\#\Gamma = 1$ , i.e., the missing case of Theorem 1.4. We say a measure  $\mu \in M(\mathbb{T}^d)$  is a *positive measure* if  $\mu(A) \geq 0$  for all Borel sets  $A \subseteq \mathbb{T}^d$ . A sequence  $\{a_m \in \mathbb{C} : m \in \mathbb{Z}^d\}$  is a *positive-definite sequence* if for all sequences  $\{b_m \in \mathbb{C} : m \in \mathbb{Z}^d\}$  of finite support, we have

$$\sum_{m, n \in \mathbb{Z}^d} a_{m-n} b_m \overline{b_n} \geq 0.$$

For  $y \in \mathbb{T}^d$ , let  $T_y : M(\mathbb{T}^d) \rightarrow M(\mathbb{T}^d)$  be the translation operator defined by  $T_y \mu(x) = \mu(x - y)$ . For  $n \in \mathbb{Z}^d$ , let  $M_n : M(\mathbb{T}^d) \rightarrow M(\mathbb{T}^d)$  be the modulation operator defined by  $M_n \mu(x) = e^{2\pi i n \cdot x} \mu(x)$ .

**Proposition 2.8.** *Let  $\mu \in M(\mathbb{T}^d)$  and let  $\Lambda \subseteq \mathbb{Z}^d$  be a finite subset. Suppose there exists  $n \in \Lambda$  such that  $(M_{-n}\mu)^\wedge|_{\Lambda-n}$  extends to a positive-definite sequence on  $\mathbb{Z}^d$ . Then,  $n \in \Gamma$ , and each positive-definite extension of  $(M_{-n}\mu)^\wedge|_{\Lambda-n}$  corresponds to a positive measure  $\nu$ , such that  $M_n \nu$  is a minimal extrapolation of  $\mu$  from  $\Lambda$ .*

*Proof.* Extend  $(M_{-n}\mu)^\wedge|_{\Lambda-n}$  to a positive-definite sequence. By Herglotz' theorem, there exists a positive measure  $\nu \in M(\mathbb{T}^d)$  such that  $\widehat{\nu} = (M_{-n}\mu)^\wedge = T_{-n}\widehat{\mu}$  on  $\Lambda - n$ . Then,  $M_n \nu$  is an extrapolation of  $\mu$  from  $\Lambda$ , which shows that

$$\|\nu\| = \|M_n \nu\| \geq \varepsilon(\mu, \Lambda).$$

For the reverse inequality, since  $\nu$  is a positive measure, we have  $\|\nu\| = \widehat{\nu}(0)$ . Then,

$$\|M_n \nu\| = \|\nu\| = \widehat{\nu}(0) = |\widehat{\nu}(0)| = |\widehat{\mu}(n)| \leq \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} \leq \varepsilon(\mu, \Lambda),$$

where the last inequality follows by Proposition 2.4. This shows that  $|\widehat{\mu}(n)| = \|\nu\| = \varepsilon$ , which proves that  $n \in \Gamma$  and  $M_n \nu$  is a minimal extrapolation of  $\mu$  from  $\Lambda$ .  $\square$

**Remark 2.9.** Beurling [Beu89b] and Esseen [Ess45] essentially proved the analogue of the Proposition for the special case that  $n = 0$ , for  $\mathbb{R}$  instead of  $\mathbb{T}^d$ . Proposition 2.8 generalizes their result to handle situations when  $0 \notin \Lambda$ . This is important because from the viewpoint of Proposition 2.12c, the super-resolution problem is invariant under simultaneous translations of  $\hat{\mu}$  and  $\Lambda$ , which means that  $0 \in \mathbb{Z}^d$  is no more special than any other point  $n \in \mathbb{Z}^d$ .

**Remark 2.10.** Proposition 2.8 suggests that the case  $\#\Gamma = 1$  is special compared to the cases  $\#\Gamma = 0$  or  $\#\Gamma \geq 2$  because, when  $\#\Gamma = 1$ , there may exist absolutely continuous minimal extrapolations. In Example 3.2,  $\#\Gamma = 1$ , and there exist uncountably many discrete and positive absolutely continuous minimal extrapolations.

**Remark 2.11.** Suppose that  $\mu \in M(\mathbb{T})$  and  $\hat{\mu}|_{\Lambda}$  can be extended to a positive-definite sequence on  $\mathbb{Z}$ . In theory, there are an infinite number of such extensions, and one particular method of choosing such an extension is called the *Maximum Entropy Method* (MEM). According to MEM, one extends  $\hat{\mu}|_{\Lambda}$  to the positive-definite sequence  $\{a_m \in \mathbb{C} : m \in \mathbb{Z}\}$  whose corresponding density function  $f \in L^1(\mathbb{T})$  is the unique maximizer of a specific logarithmic integral associated with the physical notion of entropy, e.g., see [Ben96, Theorems 3.6.3 and 3.6.6]. MEM is related to spectral estimation methods [Chi78], the maximum likelihood method [Chi78], and moment problems [Lan87].

**2.6. Basic properties of minimal extrapolation.** Our next goal is to examine the symmetries of the minimal extrapolations. We are interested in the vector space operations, namely, addition of measures and the multiplication of measures by complex constants. We are also interested in the operations that are well-behaved under the Fourier transform on  $\mathbb{T}^d$ , namely, translation on the torus, modulation by integers, convolution of measures, and product of measures.

**Proposition 2.12.** *Let  $\mu \in M(\mathbb{T}^d)$ ,  $\Lambda \subseteq \mathbb{Z}^d$  be a finite subset,  $c \in \mathbb{C}$  non-zero, and  $y \in \mathbb{T}^d$ .*

- (a) *Multiplication by constants is bijective:  $c \varepsilon(\mu, \Lambda) = \varepsilon(c\mu, \Lambda)$ , and  $\nu \in \mathcal{E}(\mu, \Lambda)$  if and only if  $c\nu \in \mathcal{E}(c\mu, \Lambda)$ .*
- (b) *Translation is bijective:  $\varepsilon(\mu, \Lambda) = \varepsilon(T_y\mu, \Lambda)$ , and  $\nu \in \mathcal{E}(\mu, \Lambda)$  if and only if  $T_y\nu \in \mathcal{E}(T_y\mu, \Lambda)$ .*
- (c) *Minimal extrapolation is invariant under simultaneous shifts of  $\hat{\mu}$  and  $\Lambda$ :  $\varepsilon(\mu, \Lambda) = \varepsilon(M_n\mu, \Lambda + n)$ , and  $\mathcal{E}(\mu, \Lambda) = \mathcal{E}(M_n\mu, \Lambda + n)$ .*
- (d) *The product of minimal extrapolations is a minimal extrapolation for the product: For  $j = 1, 2$ , let  $\mu_j \in M(\mathbb{T}^{d_j})$ , let  $\Lambda_j \subseteq \mathbb{Z}^{d_j}$  be a finite subset, and let  $\nu_j \in \mathcal{E}(\mu_j, \Lambda_j)$ . Then  $\varepsilon(\mu_1, \Lambda_1)\varepsilon(\mu_2, \Lambda_2) = \varepsilon(\mu_1 \times \mu_2, \Lambda_1 \times \Lambda_2)$ , and  $\nu_1 \times \nu_2 \in \mathcal{E}(\mu_1 \times \mu_2, \Lambda_1 \times \Lambda_2)$ .*

*Proof.*

- (a) If  $\nu \in \mathcal{E}(\mu, \Lambda)$ , then  $c\nu$  is an extrapolation of  $c\mu$  from  $\Lambda$ . Suppose  $c\nu \notin \mathcal{E}(c\mu, \Lambda)$ . Then there exists  $\sigma$  such that  $\hat{\sigma} = c\hat{\mu}$  on  $\Lambda$  and  $\|\sigma\| < \|c\nu\|$ . Thus,  $\hat{\sigma}/c = \hat{\mu}$  on  $\Lambda$  and  $\|\sigma/c\| < \|\nu\|$ , and this contradicts the assumption that  $\nu \in \mathcal{E}(\mu, \Lambda)$ . The converse follows by a similar argument.
- (b) If  $\nu \in \mathcal{E}(\mu, \Lambda)$ , then  $T_y\nu$  is an extrapolation of  $T_y\mu$  from  $\Lambda$ . Suppose  $T_y\nu \notin \mathcal{E}(T_y\mu, \Lambda)$ . Then there exists  $\sigma$  such that  $\hat{\sigma} = (T_y\mu)^\wedge$  on  $\Lambda$  and  $\|\sigma\| < \|T_y\nu\| = \|\nu\|$ . Then  $(T_{-y}\sigma)^\wedge = \hat{\mu}$  on  $\Lambda$  and  $\|T_{-y}\sigma\| = \|\sigma\| < \|\nu\|$ , which contradicts the assumption that  $\nu \in \mathcal{E}(\mu, \Lambda)$ . The converse follows by a similar argument.

(c) If  $f \in U(\mathbb{T}^d; \Lambda)$ , then  $M_n f \in U(\mathbb{T}^d; \Lambda + n)$ . By Parseval's theorem,

$$\langle f, \mu \rangle = \sum_{m \in \Lambda} \widehat{f}(m) \overline{\widehat{\mu}(m)} = \sum_{m \in \Lambda + n} (M_n f)^\wedge(m) \overline{(M_n \mu)^\wedge(m)} = \langle M_n f, M_n \mu \rangle.$$

Using Proposition 2.1e, we see that  $\varepsilon(\mu, \Lambda) = \varepsilon(M_n \mu, \Lambda + n)$ .

If  $\nu \in \mathcal{E}(\mu, \Lambda)$ , then  $M_n \nu$  is an extrapolation of  $M_n \mu$  from  $\Lambda + n$ , and  $\|M_n \nu\| = \|\nu\| = \varepsilon(\mu, \Lambda) = \varepsilon(M_n \mu, \Lambda + n)$ . The converse follows similarly.

(d) For convenience, let  $\mu = \mu_1 \times \mu_2$ ,  $\nu = \nu_1 \times \nu_2$ ,  $\Lambda = \Lambda_1 \times \Lambda_2$ ,  $\varepsilon_j = \varepsilon(\mu_j, \Lambda_j)$ , and  $\varepsilon = \varepsilon(\mu, \Lambda)$ . Since  $\nu$  is an extrapolation of  $\mu$ , by Proposition 2.4, we have

$$\varepsilon \leq \|\nu\| = \|\nu_1\| \|\nu_2\| = \varepsilon_1 \varepsilon_2.$$

To see the reverse inequality, we use Proposition 2.1f, and see that there exist  $\varphi_j \in U(\mathbb{T}^{d_j}; \Lambda_j)$  such that  $\varepsilon_j = \langle \varphi_j, \mu_j \rangle$ , for  $j = 1, 2$ . Let  $\varphi = \varphi_1 \otimes \varphi_2$  and observe that  $\varphi \in U(\mathbb{T}^d; \Lambda)$ . By Proposition 2.1e, we have

$$\varepsilon = \max_{f \in U(\mathbb{T}^d; \Lambda)} |\langle f, \mu \rangle| \geq |\langle \varphi, \mu \rangle| \geq \langle \varphi, \mu \rangle = \langle \varphi_1, \mu_1 \rangle \langle \varphi_2, \mu_2 \rangle = \varepsilon_1 \varepsilon_2.$$

This shows that  $\varepsilon = \varepsilon_1 \varepsilon_2$ , and, since  $\|\nu\| = \varepsilon_1 \varepsilon_2$ , we conclude that  $\nu \in \mathcal{E}(\mu, \Lambda)$ . □

While minimal extrapolation is well-behaved under translation, it is not well behaved under modulation. This is because the Fourier transform of modulation is translation, and so  $\widehat{\mu}|_\Lambda$  and  $(M_n \mu)^\wedge|_\Lambda$  are, in general, not equal. In contrast, the Fourier transform of translation is modulation, and so  $\widehat{\mu}|_\Lambda$  and  $(T_y \mu)^\wedge|_\Lambda$  only differ by a phase factor. We shall prove these statements in Proposition 2.13.

**Proposition 2.13.**

- (a) For  $j = 1, 2$ , there exist  $\mu_j \in M(\mathbb{T})$ , a finite subset  $\Lambda \subseteq \mathbb{Z}$ , and  $\nu_j \in \mathcal{E}(\mu_j, \Lambda)$ , such that  $\nu_1 + \nu_2 \notin \mathcal{E}(\mu_1 + \mu_2, \Lambda)$ .
- (b) There exist  $\mu \in M(\mathbb{T})$ , a finite subset  $\Lambda \subseteq \mathbb{Z}$ ,  $\nu \in \mathcal{E}(\mu, \Lambda)$ , and  $n \in \mathbb{Z}$ , such that  $M_n \nu \notin \mathcal{E}(M_n \mu, \Lambda)$ .
- (c) For  $j = 1, 2$ , there exist  $\mu_j \in M(\mathbb{T})$ , a finite subset  $\Lambda \subseteq \mathbb{Z}$ , and  $\nu_j \in \mathcal{E}(\mu_j, \Lambda)$ , such that  $\nu_1 * \nu_2 \notin \mathcal{E}(\mu_1 * \mu_2, \Lambda)$ .

*Proof.*

- (a) Let  $\mu_1 = \delta_0 + \delta_{1/2}$ ,  $\mu_2 = -\delta_0 - \delta_{1/2}$ , and  $\Lambda = \{-1, 0, 1\}$ . By Example 3.2, we have  $\nu_1 = \delta_0 + \delta_{1/2} \in \mathcal{E}(\mu_1, \Lambda)$ , and  $\nu_2 = -\delta_{1/4} - \delta_{3/4} \in \mathcal{E}(\mu_2, \Lambda)$ . Then,  $\mu_1 + \mu_2 = 0$ , and so  $\varepsilon(\mu_1 + \mu_2, \Lambda) = 0$ . However,  $\nu_1 + \nu_2 \notin \mathcal{E}(\mu_1 + \mu_2, \Lambda)$  because  $\|\nu_1 + \nu_2\| = \|\delta_0 - \delta_{1/4} + \delta_{1/2} - \delta_{3/4}\| > 0$ .
- (b) Let  $\mu = \delta_0 + \delta_{1/2}$ ,  $\Lambda = \{-1, 0, 1\}$ , and  $n = -1$ . By Example 3.2, we have  $\nu = \delta_{1/4} + \delta_{3/2} \in \mathcal{E}(\mu, \Lambda)$ . However,  $M_{-1} \mu = \delta_0 - \delta_{1/2}$ , and by Example 3.3,  $\mathcal{E}(M_{-1} \mu, \Lambda) = \{\delta_0 - \delta_{1/2}\}$ . Thus,  $M_{-1} \nu \notin \mathcal{E}(M_{-1} \mu, \Lambda)$ .
- (c) Let  $\mu_1 = \delta_0 + \delta_{1/2}$ ,  $\mu_2 = \delta_0 - \delta_{1/2}$ , and  $\Lambda = \{-1, 0, 1\}$ . Then  $(\mu_1 * \mu_2)^\wedge = 0$  on  $\Lambda$ , which implies  $\varepsilon(\mu_1 * \mu_2, \Lambda) = 0$ . By Examples 3.2 and 3.3,  $\nu_1 = \mu_1 \in \mathcal{E}(\mu_1, \Lambda)$  and  $\nu_2 = \mu_2 \in \mathcal{E}(\mu_2, \Lambda)$ . However,  $\nu_1 * \nu_2 \notin \mathcal{E}(\mu_1 * \mu_2, \Lambda)$  because  $\|\nu_1 * \nu_2\| = \|\delta_0 - \delta_1\| > 0$ . □

## 3. EXAMPLES

**3.1. Discrete measures.** There are several reasons why we are interested in computing the minimal extrapolations of discrete measures. They are the simplest types of measures, and so, their minimal extrapolations can be computed rather easily. By Theorem 1.4, the minimal extrapolations of a non-discrete measure are sometimes discrete measures, so they appear naturally in our analysis. As discussed in Section 1.4, examples of  $\mu$  that have minimal extrapolations supported in a lattice can be interpreted in the context of deterministic compressed sensing. Examples 3.2-3.5 and Example 3.7 can be written in the context of compressed sensing.

**Remark 3.1.** In view of Proposition 2.12 a,b, and without loss of generality, we can assume any discrete measure  $\mu = \sum_{k=1}^{\infty} a_k \delta_{x_k} \in M(\mathbb{T}^d)$ , where  $\sum_{k=1}^{\infty} |a_k| < \infty$ , can be written as  $\mu = \delta_0 + \sum_{k=2}^{\infty} a'_k \delta_{x'_k} \in M(\mathbb{T}^d)$ , where  $\sum_{k=2}^{\infty} |a'_k| < \infty$ .

**Example 3.2.** Let  $\mu = \delta_0 + \delta_{1/2}$ , and  $\Lambda = \{-1, 0, 1\}$ . We have  $\widehat{\mu}(0) = 2$ , and  $\widehat{\mu}(\pm 1) = 0$ . Clearly  $\|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$ . By Proposition 2.4,  $\varepsilon = \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$ , which implies  $\mu \in \mathcal{E}$ .

Further, there is an uncountable number of discrete minimal extrapolations. To see this, for each  $y \in \mathbb{T}$  and any integer  $K \geq 2$ , define

$$\nu_{y,K} = \frac{2}{K} \sum_{k=0}^{K-1} \delta_{y + \frac{k}{K}}.$$

A straightforward calculation shows that  $\nu_{y,K}$  is an extrapolation and that  $\|\nu_{y,K}\| = \varepsilon$ .

Also, we can construct positive absolutely continuous minimal extrapolations. One example is the constant function  $f \equiv 2$  on  $\mathbb{T}$ . For other examples, let  $N \geq 2$  and let  $F_N \in C^\infty(\mathbb{T})$  be the Fejér kernel,

$$F_N(x) = \sum_{n=-N}^N \left(1 - \frac{|n|}{N+1}\right) e^{2\pi i n x}.$$

For any  $c > 0$  such that  $c \leq (2N+2)/(3N+1)$ , extend  $\widehat{\mu}|_\Lambda$  to the sequence  $\{(a_{N,c})_m : m \in \mathbb{Z}\}$ , where

$$(a_{N,c})_m = \begin{cases} 2 & m = 0, \\ c \left(1 - \frac{|m|}{N+1}\right) & 2 \leq |m| \leq N, \\ 0 & \text{otherwise.} \end{cases}$$

Consider the real-valued function

$$\begin{aligned} f_{N,c}(x) &= 2 + \sum_{n=-N}^{-2} (a_{N,c})_n e^{2\pi i n x} + \sum_{n=2}^N (a_{N,c})_n e^{2\pi i n x} \\ &= 2 + c \sum_{n=-N}^{-2} \left(1 - \frac{|n|}{N+1}\right) e^{2\pi i n x} + c \sum_{n=2}^N \left(1 - \frac{|n|}{N+1}\right) e^{2\pi i n x}. \end{aligned}$$

We check that  $\widehat{f_{N,c}}(m) = (a_{N,c})_m$  for all  $m \in \mathbb{Z}$ , which implies  $f_{N,c}$  is an extrapolation of  $\mu$ . Using the upper bound on  $c$ , we have, for all  $x \in \mathbb{T}$ , that

$$2 \geq c + 2c \left(1 - \frac{1}{N+1}\right) \cos(2\pi x) = c + c \left(1 - \frac{1}{N+1}\right) e^{2\pi i x} + c \left(1 - \frac{1}{N+1}\right) e^{-2\pi i x}.$$



Using this inequality, and the definitions of  $F_N$  and  $f_{N,c}$ , we have

$$f_{N,c}(x) \geq cF_N \geq 0.$$

Since  $f_{N,c} \geq 0$ , we also have

$$\|f_{N,c}\|_1 = \int_{\mathbb{T}} f_{N,c}(x) dx = 2 + \int_{\mathbb{T}} \left( \sum_{n=-N}^{-2} (a_{N,c})_n e^{2\pi i n x} + \sum_{n=2}^N (a_{N,c})_n e^{2\pi i n x} \right) dx = 2 = \varepsilon.$$

Thus, for any  $N \geq 2$  and  $c \leq (2N+2)/(3N+1)$ ,  $f_{N,c}$  is a *positive absolutely continuous* minimal extrapolation. Hence, we have constructed an uncountable number of positive absolutely continuous minimal extrapolations.

**Example 3.3.** Let  $\mu = \delta_0 - \delta_{1/2}$ , and  $\Lambda = \{-1, 0, 1\}$ . We have  $\hat{\mu}(0) = 0$ , and  $\hat{\mu}(\pm 1) = 2$ . Further, we have  $\|\hat{\mu}\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$ , so that by Proposition 2.4, we have  $\varepsilon = \|\hat{\mu}\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$  and  $\mu \in \mathcal{E}$ .

Consequently,  $\Gamma = \{-1, 1\}$ . By Theorem 1.4b, there exists  $\alpha_{-1,1} \in \mathbb{R}/\mathbb{Z}$  satisfying

$$e^{2\pi i \alpha_{-1,1}} = \frac{\hat{\mu}(-1)}{\hat{\mu}(1)} = 1,$$

and the minimal extrapolations are supported in the set  $\{x \in \mathbb{T} : 2x \in \mathbb{Z}\} = \{0, 1/2\}$ . This implies each  $\nu \in \mathcal{E}$  is discrete and can be written as  $\nu = a_1 \delta_0 + a_2 \delta_{1/2}$ . In theory,  $a_1, a_2$  depend on  $\nu$ , so we cannot conclude uniqueness yet.

The matrix  $E$  from (2.4) is

$$E(-1, 0, 1; 0, 1/2) = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Clearly,  $E$  has full column rank, so, by Proposition 2.7,  $\mu$  is the unique minimal extrapolation. Thus, super-resolution reconstruction of  $\mu$  from  $\Lambda$  is possible.

**Example 3.4.** Let  $\mu = \delta_0 - \delta_{1/4}$ , and let  $\Lambda = \{-1, 0, 1\}$ . We have  $\hat{\mu}(\pm 1) = 1 \pm i = \sqrt{2}e^{\pm \pi i/4}$ , and  $\hat{\mu}(0) = 0$ . Note that  $\|\hat{\mu}\|_{\ell^\infty(\Lambda)} = \sqrt{2} < 2 = \|\mu\|$ , which shows that  $\sqrt{2} \leq \varepsilon \leq 2$ . We claim that  $\varepsilon = \sqrt{2}$ . To see this, consider  $\nu = (-\delta_{3/8} + \delta_{7/8})/\sqrt{2}$ . We verify that  $\|\nu\| = \sqrt{2}$  and that  $\nu$  is an extrapolation. By Proposition 2.4,  $\varepsilon = \sqrt{2}$  and  $\nu \in \mathcal{E}$ . This also implies  $\mu \notin \mathcal{E}$ , and so super-resolution reconstruction of  $\mu$  from  $\Lambda$  is impossible.

We claim that  $\nu$  is the unique minimal extrapolation. The matrix  $E$  from (2.4) is

$$E(-1, 0, 1; 3/8, 7/8) = \begin{pmatrix} e^{2\pi i 3/8} & e^{2\pi i 7/8} \\ 1 & 1 \\ e^{-2\pi i 3/8} & e^{-2\pi i 7/8} \end{pmatrix},$$

which we observe to have full column rank. By Proposition 2.7, we conclude that  $\nu$  is the unique minimal extrapolation. Thus, super-resolution reconstruction of  $\nu$  from  $\Lambda$  is possible.

We explain the derivation of  $\nu$ . We guess that  $\varepsilon = \sqrt{2}$  and see what Theorem 1.4b implies. Under this assumption that  $\varepsilon = \sqrt{2}$ , we have  $\Gamma = \{-1, 1\}$ . By Theorem 1.4b, there exists  $\alpha_{-1,1} \in \mathbb{R}/\mathbb{Z}$  satisfying

$$e^{2\pi i \alpha_{-1,-1}} = \frac{\hat{\mu}(1)}{\hat{\mu}(-1)} = e^{\pi i/2},$$

and the minimal extrapolations are supported in  $\{x \in \mathbb{T} : 2x + 1/4 \in \mathbb{Z}\} = \{3/8, 7/8\}$ . Hence, if  $\varepsilon = \sqrt{2}$ , then every  $\sigma \in \mathcal{E}$  is of the form  $\sigma = a_1 \delta_{3/8} + a_2 \delta_{7/8}$ . Thus, by definition of a minimal extrapolation,  $\|\sigma\| = \sqrt{2}$  and  $\hat{\mu} = \hat{\sigma}$  on  $\Lambda$ . Using this information, we solve for the coefficients  $a_1, a_2$ , and compute that  $|a_1| = |a_2| = \sqrt{2}$ ,  $a_1 = -a_2$ , and  $a_1 = -\sqrt{2}/2$ . Thus, we obtain that  $\sigma = (-\delta_{3/8} + \delta_{7/8})/\sqrt{2}$ . From here, we simply check that  $\nu = \sigma$  is, in fact, a minimal extrapolation.

**Example 3.5.** Let  $\mu = \delta_0 + e^{\pi i/3} \delta_{1/3}$  and let  $\Lambda = \{-1, 0, 1\}$ . We have  $\hat{\mu}(-1) = 0$ ,  $\hat{\mu}(0) = 1 + e^{\pi i/3} = \sqrt{3} e^{\pi i/6}$ , and  $\hat{\mu}(1) = 1 + e^{-\pi i/3} = \sqrt{3} e^{-\pi i/6}$ .

Suppose, for the purpose of obtaining a contradiction, that  $\varepsilon = \|\hat{\mu}\|_{\ell^\infty(\Lambda)} = \sqrt{3}$ . Then  $\Gamma = \{0, 1\}$ . By Theorem 1.4b, there is  $\alpha_{0,1} \in \mathbb{R}/\mathbb{Z}$  such that

$$e^{2\pi i \alpha_{0,1}} = \frac{\hat{\mu}(0)}{\hat{\mu}(1)} = e^{\pi i/3},$$

and each  $\nu \in \mathcal{E}$  is of the form  $\nu = a \delta_{1/6}$  for some  $a \in \mathbb{C}$ . Then,  $|\hat{\nu}| = |a|$  on  $\mathbb{Z}$  and, in particular,  $\hat{\mu} \neq \hat{\nu}$  on  $\Lambda$ , which is a contradiction.

Thus,  $\varepsilon > \sqrt{3}$ , i.e.,  $\Gamma = \emptyset$ . Therefore, Theorem 1.4a applies, and so there is a finite set  $S$  such that  $\text{supp}(\nu) \subseteq S$  for each  $\nu \in \mathcal{E}$ . In particular, each  $\nu \in \mathcal{E}$  is discrete. Hence, we have to solve the optimization problem given in Proposition 2.1e, which is

$$\varepsilon = \max \left\{ \left| a\sqrt{3} e^{\pi i/6} + b\sqrt{3} e^{-\pi i/6} \right| : \forall x \in \mathbb{T}, |ae^{2\pi i x} + b + ce^{-2\pi i x}| \leq 1, a, b, c \in \mathbb{C} \right\}.$$

This optimization problem can be written as a semi-definite program, see [CFG14, Corollary 4.1, page 936]. After obtaining numerical approximations to the optimizers of this problem, we guess that the exact optimizers are

$$a = \frac{2}{3\sqrt{3}} e^{-\pi i/6}, \quad b = \frac{4}{3\sqrt{3}} e^{\pi i/6}, \quad c = -\frac{i}{3\sqrt{3}}.$$

These values of  $a, b, c$  are, in fact, the optimizers because  $a\sqrt{3} e^{\pi i/6} + b\sqrt{3} e^{-\pi i/6} = 2$  and  $|ae^{2\pi i x} + b + ce^{-2\pi i x}| \leq 1$  for all  $x \in \mathbb{T}$ . Thus,  $\varepsilon = 2$  and  $\mu \in \mathcal{E}$ . Since  $|ae^{2\pi i x} + b + ce^{-2\pi i x}| = 1$  precisely on  $S = \{0, 1/3\}$ , by Theorem 1.4a, the minimal extrapolations are supported in  $S$ .

The matrix  $E$  from (2.4) is

$$E(-1, 0, 1; 0, 1/3) = \begin{pmatrix} 1 & e^{-2\pi i/3} \\ 1 & 1 \\ 1 & e^{2\pi i/3} \end{pmatrix}.$$

Since  $E$  has full rank, by Proposition 2.7,  $\mu$  is the unique minimal extrapolation. Thus, super-resolution reconstruction of  $\mu$  from  $\Lambda$  is possible.

The following example illustrates that if  $\mu$  is a sum of two Dirac measures, then their supports have to be sufficiently spaced apart in order for super-resolution of  $\mu$  to be possible. This shows that, in general, a minimum separation condition is necessary to super-resolve a sum of two Dirac measures, see [CFG14].

**Example 3.6.** Let  $\mu_y = \delta_0 - \delta_y$  for some non-zero  $y \in \mathbb{T}^d$  and let  $\Lambda \subseteq \mathbb{Z}^d$  be a finite subset. We claim that if  $y$  is sufficiently small depending on  $\Lambda$ , then  $\mu_y \notin \mathcal{E}(\mu_y, \Lambda)$ . Note that  $\|\mu_y\| = 2$  for any  $y \in \mathbb{T}^d$ . Let  $\eta$  denote the normalized Lebesgue measure on  $\mathbb{T}^d$  and

define the measures  $\nu_y$  by the formula

$$\nu_y(x) = \sum_{m \in \Lambda} \widehat{\mu}_y(m) e^{2\pi i m \cdot x} \eta(x).$$

By construction,  $\nu_y$  is an extrapolation of  $\mu_y$  because, for each  $n \in \Lambda$ ,

$$\widehat{\nu}_y(n) = \int_{\mathbb{T}^d} e^{-2\pi i n \cdot x} d\nu_y(x) = \sum_{m \in \Lambda} \widehat{\mu}_y(m) \int_{\mathbb{T}^d} e^{-2\pi i (n-m) \cdot x} dx = \widehat{\mu}_y(n).$$

Then, as  $y \rightarrow 0$ ,

$$\|\nu_y\| = \int_{\mathbb{T}^d} \left| \sum_{m \in \Lambda} \widehat{\mu}_y(m) e^{2\pi i m \cdot x} \right| dx = \int_{\mathbb{T}^d} \left| \sum_{m \in \Lambda} (1 - e^{-2\pi i m \cdot y}) e^{2\pi i m \cdot x} \right| dx \rightarrow 0.$$

Thus, we take  $y$  sufficiently small so that  $\|\nu_y\| < 2 = \|\mu_y\|$ , and, hence,  $\mu_y \notin \mathcal{E}(\mu_y, \Lambda)$ . Note that this argument does not contradict Proposition 2.4 because  $\|\widehat{\mu}_y\|_{\ell^\infty(\Lambda)} \rightarrow 0$  as  $y \rightarrow 0$ . Thus, for  $y$  sufficiently small, super-resolution reconstruction of  $\mu_y$  from  $\Lambda$  is impossible.

**3.2. Singular continuous measures.** The following example is an analogue of Example 3.2 for higher dimensions.

**Example 3.7.** Let  $\mu = \delta_{(0,0)} + \delta_{(1/2,1/2)}$ , and  $\Lambda = \{-1, 0, 1\}^2 \setminus \{(1, -1), (-1, 1)\}$ . Then,  $\widehat{\mu}(m) = 1 + e^{-\pi i(m_1+m_2)}$ , and, in particular,  $\widehat{\mu}(1, 1) = \widehat{\mu}(-1, -1) = \widehat{\mu}(0, 0) = 2$  and  $\widehat{\mu}(\pm 1, 0) = \widehat{\mu}(0, \pm 1) = 0$ . We deduce that  $\varepsilon = \|\mu\| = \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = 2$  from Proposition 2.4, and so  $\mu \in \mathcal{E}$ .

Further,  $\Gamma = \{(0, 0), (1, 1), (-1, -1)\}$ , and so  $\#\Gamma = 3$ . According to the definition of  $\alpha_{m,n}$  in Theorem 1.4b, set  $\alpha_{(-1,-1),(0,0)} = \alpha_{(0,0),(1,1)} = \alpha_{(-1,-1),(1,1)} = 0$ . By the conclusion of Theorem 1.4b, the minimal extrapolations are supported in the set  $S = S_{(-1,-1),(0,0)} \cap S_{(0,0),(1,1)} \cap S_{(-1,-1),(1,1)}$ , where

$$\begin{aligned} S_{(-1,-1),(0,0)} &= \{x \in \mathbb{T}^2 : x \cdot (-1, -1) \in \mathbb{Z}\} = \{x \in \mathbb{T}^2 : x_1 + x_2 \in \mathbb{Z}\}, \\ S_{(0,0),(1,1)} &= \{x \in \mathbb{T}^2 : x \cdot (-1, -1) \in \mathbb{Z}\} = \{x \in \mathbb{T}^2 : x_1 + x_2 \in \mathbb{Z}\}, \\ S_{(-1,-1),(1,1)} &= \{x \in \mathbb{T}^2 : x \cdot (-2, -2) \in \mathbb{Z}\} = \{x \in \mathbb{T}^2 : 2x_1 + 2x_2 \in \mathbb{Z}\}. \end{aligned}$$

It follows that the minimal extrapolations are supported in

$$S = S_{(-1,-1),(0,0)} \cap S_{(0,0),(1,1)} \cap S_{(-1,-1),(1,1)} = \{x \in \mathbb{T}^2 : x_1 + x_2 = 1\}.$$

We can construct other discrete minimal extrapolations besides  $\mu$ . For each  $y \in \mathbb{T}$  and for each integer  $K \geq 2$ , define the measure

$$\nu_{y,K} = \frac{2}{K} \sum_{k=0}^{K-1} \delta_{(y+\frac{k}{K}, 1-y-\frac{k}{K})}.$$

We claim  $\nu_{y,K}$  is a minimal extrapolation. We have  $\|\nu_{y,K}\| = \varepsilon$ , and

$$\widehat{\nu_{y,K}}(m) = e^{-2\pi i(m_1 y - m_2 y)} e^{-2\pi i m_2} \frac{2}{K} \sum_{k=0}^{K-1} e^{-2\pi i(m_1 - m_2)k/K}.$$

We see that  $\widehat{\nu_{y,K}} = \widehat{\mu}$  on  $\Lambda$ , which proves the claim.

We can also construct continuous singular minimal extrapolations. Let  $\sigma = \sqrt{2}\sigma_S$ , where  $\sigma_S$  is the *surface measure* of the Borel set  $S$ . We readily verify that  $\|\sigma\| = \varepsilon$  and

$$\widehat{\sigma}(m) = \sqrt{2} \int_{\mathbb{T}^2} e^{-2\pi i m \cdot x} d\sigma_S = 2e^{-2\pi i m_2} \int_0^1 e^{-2\pi i(m_1 - m_2)t} dt = 2\delta_{m_1, m_2},$$

which proves that  $\sigma \in \mathcal{E}$ . In particular,  $S$  is the smallest set that contains the support of all the minimal extrapolations.

Since  $\mu$  is not the unique minimal extrapolation, super-resolution reconstruction of  $\mu$  from  $\Lambda$  is impossible.

**Example 3.8.** For an integer  $q \geq 3$ , let  $C_q$  be the *middle  $1/q$ -Cantor set*, which is defined by  $C_q = \bigcap_{k=0}^{\infty} C_{q,k}$ , where  $C_{q,0} = [0, 1]$  and

$$C_{q,k+1} = \frac{C_{q,k}}{q} \cup \left( (1 - q) + \frac{C_{q,k}}{q} \right).$$

Let  $F_q: [0, 1] \rightarrow [0, 1]$  be the *Cantor-Lebesgue* function on  $C_q$ , which is defined by the point-wise limit of the sequence  $\{F_{q,k}\}$ , where  $F_{q,0}(x) = x$  and

$$F_{q,k+1}(x) = \begin{cases} \frac{1}{2}F_{q,k}(qx) & 0 \leq x \leq \frac{1}{q}, \\ \frac{1}{2} & \frac{1}{q} \leq x \leq \frac{q-1}{q}, \\ \frac{1}{2}F_{q,k}(qx - (q-1)) + \frac{1}{2} & \frac{q-1}{q} \leq x \leq 1. \end{cases}$$

By construction,  $F_q(0) = 0$ ,  $F_q(1) = 1$ , and  $F_q$  is non-decreasing and uniformly continuous on  $[0, 1]$ . Thus,  $F_q$  can be uniquely identified with the measure  $\sigma_q \in M(\mathbb{T})$ , and  $\|\sigma_q\| = 1$ . Since  $F'_q = 0$  a.e. and  $F_q$  does not have any jump discontinuities,  $\sigma_q$  is a continuous singular measure, with zero discrete part. The Fourier coefficients of  $\sigma_q$  are

$$\widehat{\sigma}_q(m) = (-1)^m \prod_{k=1}^{\infty} \cos(\pi m q^{-k}(1 - q)),$$

see [Zyg59, pages 195-196]. In particular, for any integer  $n \geq 1$ , we have

$$\widehat{\sigma}_q(q^n) = (-1)^{q^n} \prod_{k=1}^{\infty} \cos(\pi q^{-k}(1 - q)),$$

which is convergent and independent of  $n$ . Since  $\widehat{\sigma}_q(0) = \|\sigma_q\| = 1$ , we immediately see that  $\varepsilon = 1$  and  $\sigma_q \in \mathcal{E}$ . Again, we cannot determine whether  $\sigma_q$  is the unique minimal extrapolation because Theorem 1.4 cannot handle the case  $\#\Gamma = 1$ , see Remark 2.10.

**Example 3.9.** Let  $\sigma_A, \sigma_B \in M(\mathbb{T}^d)$  be the surface measures of the Borel sets  $A = \{x \in \mathbb{T}^2: x_2 = 0\}$  and  $B = \{x \in \mathbb{T}^2: x_2 = 1/2\}$ , respectively. Let  $\mu = \sigma_A + \sigma_B$ , and  $\Lambda = \{-2, -1, \dots, 1\}^2$ . Then,

$$\widehat{\mu}(m) = \int_0^1 e^{2\pi i m_1 t} dt + \int_0^1 e^{2\pi i(m_1 t + m_2/2)} dt = \delta_{m_1, 0} + (-1)^{m_2} \delta_{m_1, 0}.$$

We immediately see that  $\varepsilon = \|\widehat{\mu}\|_{\ell^\infty(\Lambda)} = \|\mu\| = 2$ , which implies  $\mu \in \mathcal{E}$ . Then,  $\Gamma = \{(0, 0), (0, 2), (0, -2)\}$ , and, by Theorem 1.4b, the minimal extrapolations are supported in  $\{x \in \mathbb{T}^2: x_2 = 0\} \cup \{x \in \mathbb{T}^2: x_2 = 1/2\}$ . Determining whether  $\mu$  is the unique minimal extrapolation is beyond the theory we have developed herein, and we shall examine this uniqueness problem in [BL16].

## 4. ACKNOWLEDGEMENTS

The first named author gratefully acknowledges the support of ARO Grant W911 NF-15-1-0112, ARO Grant W911 NF-16-1-0008, and DTRA Grant 1-13-1-0015.

## REFERENCES

- [ADCG15] Jean-Marc Azaïs, Yohann De Castro, and Fabrice Gamboa. Spike detection from inaccurate samplings. *Applied and Computational Harmonic Analysis*, 38(2):177–195, 2015.
- [ASB14] Céline Aubel, David Stotz, and Helmut Bölcskei. Super-resolution from short-time Fourier transform measurements. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 36–40. IEEE, 2014.
- [ASB15] Céline Aubel, David Stotz, and Helmut Bölcskei. A theory of super-resolution from short-time Fourier transform measurements. *arXiv preprint arXiv:1509.01047*, 2015.
- [AYB15] Enrico Au-Yeung and John J. Benedetto. Generalized Fourier frames in terms of balayage. *Journal of Fourier Analysis and Applications*, 21(3):472–508, 2015.
- [BC09] John J. Benedetto and Wojciech Czaaja. *Integration and Modern Analysis*. Birkhäuser Boston, Inc., Boston, 2009.
- [Ben75] John J. Benedetto. *Spectral Synthesis*. Academic Press, Inc., New York-London, 1975.
- [Ben96] John J. Benedetto. *Harmonic Analysis and Applications*. CRC Press Inc., 1996.
- [Beu89a] Arne Beurling. Balayage of Fourier-Stieltjes transforms. *The Collected Works of Arne Beurling*, 2:341–350, 1989.
- [Beu89b] Arne Beurling. Interpolation for an interval in  $\mathbb{R}^1$ . *The Collected Works of Arne Beurling*, 2:351–365, 1989.
- [BL16] John J. Benedetto and Weilin Li. Super-resolution of singular continuous measures on  $\mathbb{T}^d$ . In *Preparation*, 2016.
- [BTR13] Badri Narayan Bhaskar, Gongguo Tang, and Benjamin Recht. Atomic norm denoising with applications to line spectral estimation. *IEEE Transactions on Signal Processing*, 61(23):5987–5999, 2013.
- [CFG13] Emmanuel J. Candès and Carlos Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.
- [CFG14] Emmanuel J. Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.
- [Chi78] Donald G. Childers. *Modern Spectrum Analysis*. IEEE Press, Piscataway, New Jersey, 1978.
- [CRPW12] Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan S. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational mathematics*, 12(6):805–849, 2012.
- [CRT06] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Information Theory*, 52(2):489–509, 2006.
- [DCG12] Yohann De Castro and Fabrice Gamboa. Exact reconstruction using Beurling minimal extrapolation. *Journal of Mathematical Analysis and Applications*, 395(1):336–354, 2012.
- [Don92] David L. Donoho. Superresolution via sparsity constraints. *SIAM Journal on Mathematical Analysis*, 23(5):1309–1331, 1992.
- [Don06] David L. Donoho. Compressed sensing. *IEEE Trans. Information Theory*, 52(4):1289–1306, 2006.
- [DP15] Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.
- [Dum07] Bogdan Dumitrescu. *Positive Trigonometric Polynomials and Signal Processing Applications*. Springer, Dordrecht, 2007.
- [Ess45] Carl-Gustav Esseen. Fourier analysis of distribution functions. a mathematical study of the Laplace-Gaussian law. *Acta Mathematica*, 77(1):1–125, 1945.
- [Ess54] Carl-Gustav Esseen. A note on Fourier-Stieltjes transforms and absolutely continuous functions. *Math. Scand*, 2:153–157, 1954.
- [FG16] Carlos Fernandez-Granda. Super-resolution of point sources via convex programming. *Information and Inference*, pages 1–53, 2016.

- [Gre09] Hayit Greenspan. Super-resolution in medical imaging. *The Computer Journal*, 52(1):43–63, 2009.
- [HR55] Edwin Hewitt and Herman Rubin. The maximum value of a Fourier-Stieltjes transform. *Math. Scand*, 3:97–102, 1955.
- [HT29] Einar Hille and J.D. Tamarkin. Remarks on a known example of a monotone continuous function. *American Mathematical Monthly*, 36(5):255–264, 1929.
- [Kah70] Jean-Pierre Kahane. *Séries de Fourier Absolument Convergentes*. Springer-Verlag, Berlin-New York, 1970.
- [KLM04] Valery Khaidukov, Evgeny Landa, and Tijmen Jan Moser. Diffraction imaging by focusing-defocusing: An outlook on seismic superresolution. *Geophysics*, 69(6):1478–1490, 2004.
- [KMM07] Aggelos K. Katsaggelos, Rafael Molina, and Javier Mateos. *Super Resolution of Images and Video*. Morgan & Claypool Publishers, 2007.
- [Lan87] Henry J. Landau. Maximum entropy and the moment problem. *Bulletin of the American Mathematical Society*, 16(1):47–77, 1987.
- [Lin12] Jari Lindberg. Mathematical concepts of optical superresolution. *Journal of Optics*, 14(8):083001, 2012.
- [PK05] K.G. Puschmann and F. Kneer. On super-resolution in astronomical imaging. *Astronomy & Astrophysics*, 436(1):373–378, 2005.
- [PPK03] Sung Cheol Park, Min Kyu Park, and Moon Gi Kang. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE*, 20(3):21–36, 2003.
- [Rie99] Fred Rieke. *Spikes: Exploring the Neural code*. MIT press, 1999.
- [Rud62] Walter Rudin. *Fourier Analysis on Groups*. John Wiley & Sons, Inc., New York, 1962.
- [Sal42] Raphaël Salem. On singular monotonic functions of the Cantor type. *J. Math. Phys*, 21:69–82, 1942.
- [Sle78] David Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty - V: The discrete case. *Bell System Technical Journal*, 57(5):1371–1430, 1978.
- [TBR15] Gongguo Tang, Badri Narayan Bhaskar, and Benjamin Recht. Near minimax line spectral estimation. *IEEE Transactions on Information Theory*, 61(1):499–512, 2015.
- [TBSR13] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed sensing off the grid. *IEEE Trans. Information Theory*, 59(11):7465–7490, 2013.
- [Zyg59] Antoni Zygmund. *Trigonometric Series*, volume 1. Cambridge University Press, 2002 (1959).

NORBERT WIENER CENTER, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MARYLAND, COLLEGE PARK, MD 20742, USA

*E-mail address:* `jjb@math.umd.edu`

NORBERT WIENER CENTER, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MARYLAND, COLLEGE PARK, MD 20742, USA

*E-mail address:* `wl298@math.umd.edu`